Characterization and Learning of Causal Graphs from Hard Interventions

Zihan Zhou[*1] , Muhammad Qasim Elahi[*2] , Murat Kocaoglu[1]

[1]Johns Hopkins University

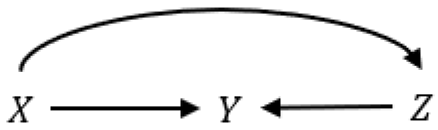
[2]Purdue University

Problem

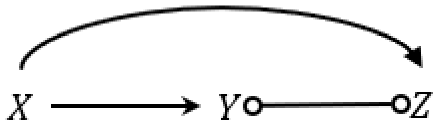
- When the underlying causal structure has latent confounders, given a tuple of hard-interventional distributions with known targets, how much can we learn about the causal graph?
- Can we characterize the interventional Markov equivalence class ($\mathcal{J} - MEC$) with a graphical representation?

Challenge

- When there is no latents, Hauser et al. show that interventions reveal the local structures.



Causal graph

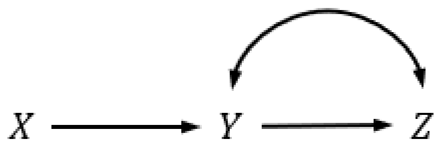


Intervene on X

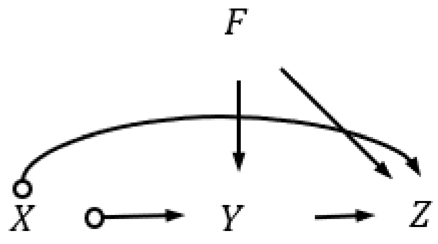
Hauser, A. and Bühlmann, P. Characterization and greedy learning of interventional markov equivalence classes of directed acyclic graphs. The Journal of Machine Learning Research, 13(1):2409–2464, 2012.

Challenge

- When the intervention is soft, Kocaoglu el al. construct augmented MAG to capture the testable separation statements.



Causal graph

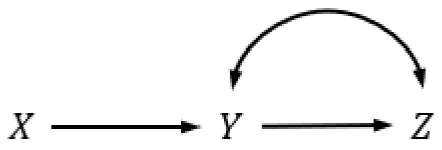


Intervene on Y

Kocaoglu, M., Jaber, A., Shanmugam, K., and Bareinboim, E. Characterization and learning of causal graphs with latent variables from soft interventions. Advances in Neural Information Processing Systems, 32, 2019.

Challenge

- Hard interventions may impose non-local separations to the graph.



Causal graph

Intervene on Y

$X \perp Z$ in the interventional domain of Y !

Do-invariance Constraints

- Given interventional data, we can test conditional independence across/within the distributions.

Do-invariance Constraints

- Given interventional data, we can test conditional independence across/within the distributions.
- Do-calculus allows the transformation across distributions.

Do-invariance Constraints

- Given interventional data, we can test conditional independence across/within the distributions.
- Do-calculus allows the transformation across distributions.
- Generalize do-calculus for arbitrary hard-interventional distributions.

Do-invariance Constraints

- Given interventional data, we can test conditional independence across/within the distributions.
- Do-calculus allows the transformation across distributions.
- Generalize do-calculus for arbitrary hard-interventional distributions.

Proposition (Generalized do-calculus for hard interventions). Let $\mathcal{D} = (\boldsymbol{V} \cup \boldsymbol{L}, \boldsymbol{E})$ be a causal graph with latents. Then, the following holds for any tuple of hard-interventional distributions $(P_{\boldsymbol{I}})_{\boldsymbol{I} \in \mathcal{I}}$ consistent with $\mathcal{D}$, where $\mathcal{I} \subseteq 2^{\boldsymbol{V}}$.

Rule 1 (conditional independence): For any $\mathbf{I} \subseteq \mathbf{V}$ *and disjoint* $\mathbf{Y}, \mathbf{Z}, \mathbf{W} \subseteq (\mathbf{V} \setminus \mathbf{I})$

$P_{\mathbf{I}}(\mathrm{y}|\mathrm{w}, \mathrm{z}) = P_{\mathbf{I}}(\mathrm{y}|\mathrm{w})$, *if* $\mathbf{Y} \perp\!\!\!\perp \mathbf{Z} | \mathbf{W}, \mathbf{I}$ *in* $\mathcal{D}_{\overline{\mathbf{I}}}$

Rule 2 (do-see): For any $\mathbf{I}, \mathbf{J} \subseteq \mathbf{V}$ *and disjoint* $\mathbf{Y}, \mathbf{W} \subseteq \mathbf{V} \setminus \mathbf{K}$, *where* $\mathbf{K} := \mathbf{I} \Delta \mathbf{J}$

$P_{\mathbf{I}}(\mathrm{y}|\mathrm{w}, \mathrm{k}) = P_{\mathbf{I},\mathbf{J}}(\mathrm{y}|\mathrm{w}, \mathrm{k}) = P_{\mathbf{J}}(\mathrm{y}|\mathrm{w}, \mathrm{k})$, *if* $(\mathbf{Y} \perp\!\!\!\perp \mathbf{K}_{\mathbf{J}} | \mathbf{W}, \mathbf{I})_{\mathcal{D}_{\overline{\mathbf{I}}, \underline{\mathbf{K}_{\mathbf{J}}}}} \wedge (\mathbf{Y} \perp\!\!\!\perp \mathbf{K}_{\mathbf{I}} | \mathbf{W}, \mathbf{J})_{\mathcal{D}_{\overline{\mathbf{J}}, \underline{\mathbf{K}_{\mathbf{I}}}}}$

Rule 3 (do-do): For any $\mathbf{I}, \mathbf{J} \subseteq \mathbf{V}$ *and disjoint* $\mathbf{Y}, \mathbf{W} \subseteq \mathbf{V} \setminus \mathbf{K}$, *where* $\mathbf{K} := \mathbf{I} \Delta \mathbf{J}$

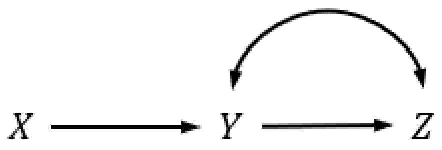
$P_{\mathbf{I}}(\mathrm{y}|\mathrm{w}) = P_{\mathbf{I},\mathbf{J}}(\mathrm{y}|\mathrm{w}) = P_{\mathbf{J}}(\mathrm{y}|\mathrm{w})$, *if* $(\mathbf{Y} \perp\!\!\!\perp \mathbf{K}_{\mathbf{J}} | \mathbf{W}, \mathbf{I})_{\mathcal{D}_{\overline{\mathbf{I}, \mathbf{K}_{\mathbf{J}}(\mathbf{W})}}} \wedge (\mathbf{Y} \perp\!\!\!\perp \mathbf{K}_{\mathbf{I}} | \mathbf{W}, \mathbf{J})_{\mathcal{D}_{\overline{\mathbf{J}, \mathbf{K}_{\mathbf{I}}(\mathbf{W})}}}$

Rule 4 (mixed do-see/do-do): For any $\mathbf{I}, \mathbf{J} \subseteq \mathbf{V}$ *and disjoint* $\mathbf{Y}, \mathbf{W} \subseteq \mathbf{V}$, *where* $\mathbf{K} := \mathbf{I} \Delta \mathbf{J}$

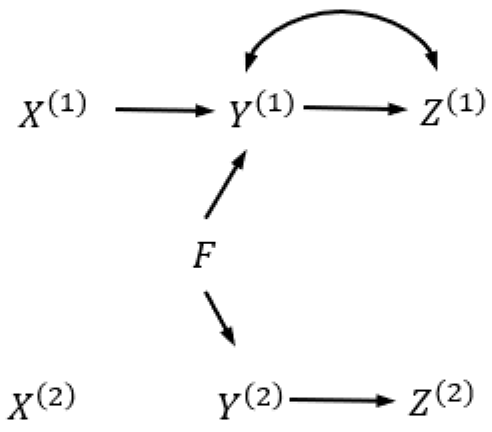
$P_{\mathbf{I}}(\mathrm{y}|\mathrm{w}) = P_{\mathbf{I},\mathbf{J}}(\mathrm{y}|\mathrm{w}, \mathrm{k}) = P_{\mathbf{J}}(\mathrm{y}|\mathrm{w})$, *if* $(\mathbf{Y} \perp\!\!\!\perp \mathbf{R}_{\mathbf{J}} | \mathbf{W}, \mathbf{I})_{\mathcal{D}_{\overline{\mathbf{I}, \mathbf{R}_{\mathbf{J}}(\mathbf{W})}}} \wedge (\mathbf{Y} \perp\!\!\!\perp \mathbf{W}_{\mathbf{J}} | \mathbf{W}, \mathbf{I})_{\mathcal{D}_{\overline{\mathbf{I}}, \underline{\mathbf{W}_{\mathbf{J}}}}} \wedge (\mathbf{Y} \perp\!\!\!\perp \mathbf{R}_{\mathbf{I}} | \mathbf{W}, \mathbf{J})_{\mathcal{D}_{\overline{\mathbf{J}, \mathbf{R}_{\mathbf{I}}(\mathbf{W})}}} \wedge (\mathbf{Y} \perp\!\!\!\perp \mathbf{W}_{\mathbf{I}} | \mathbf{W}, \mathbf{J})_{\mathcal{D}_{\overline{\mathbf{J}}, \underline{\mathbf{W}_{\mathbf{I}}}}}$

$\mathcal{J}$-Markov Equilvalence Class

- We construct the augmented pair graph to capture invariance across domains.



Causal graph with P_{obs} and P_Y

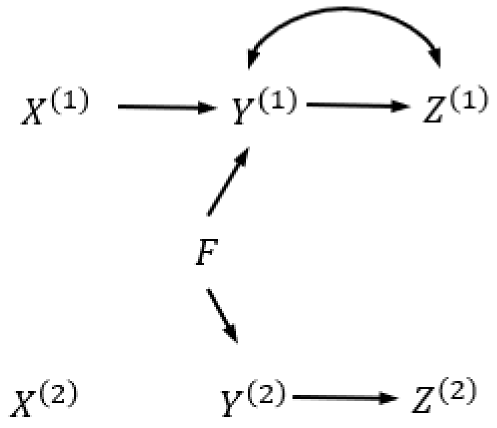


F pointing to the symmetric difference

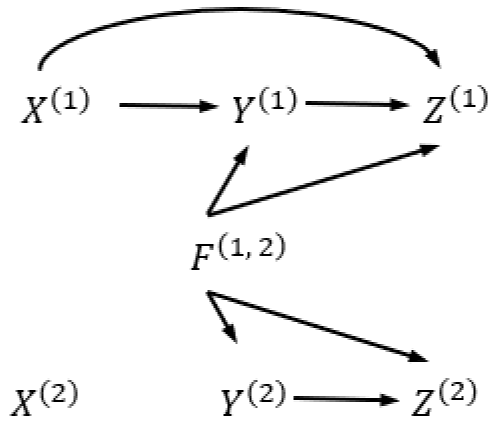
Augmented pair graph

$\mathcal{J}$-Markov Equilvalence Class

- Based on the augmented pair graphs, we build the twin augmented MAG.



Augmented pair graph



Twin augmented MAG

Make the adjacency of F symmetric

$\mathcal{I}$-Markov Equilvalence Class

- Characterize the $\mathcal{I}$-MEC based on the twin augmented MAGs.

Theorem: Given two causal graphs $\mathcal{D}_1 = (\boldsymbol{V} \cup \boldsymbol{L_1}, \boldsymbol{E_1})$ and $\mathcal{D}_2 = (\boldsymbol{V} \cup \boldsymbol{L_2}, \boldsymbol{E_2})$, and a set of intervention targets $\mathcal{I} \subseteq 2^{\boldsymbol{V}}$, $\mathcal{D}_1$ and $\mathcal{D}_2$ are $\mathcal{I}$-Markov equivalent with respect to $\mathcal{I}$ if and only if for each pair of interventions $\boldsymbol{I}, \boldsymbol{J} \in \mathcal{I}$, $\mathcal{M}_1 = Twin_{(\boldsymbol{I},\boldsymbol{J})}(\mathcal{D}_1), \mathcal{M}_2 = Twin_{(\boldsymbol{I},\boldsymbol{J})}(\mathcal{D}_2)$**:**

1. $\mathcal{M}_1$ and $\mathcal{M}_2$ have the same skeleton;
2. $\mathcal{M}_1$ and $\mathcal{M}_2$ have the same unshielded colliders;
3. If a path p is a discriminating path for a node Y in both $\mathcal{M}_1$ and $\mathcal{M}_2$, then Y is a collider on the path if and only if it is a collider on the path in the other.

$\mathcal{J}$-Augmented MAGs

- Combine twin augmented MAGs to domain centric graphs.

$\mathcal{J}$-Augmented MAGs

- Combine twin augmented MAGs to domain centric graphs.
- The same conditions for $\mathcal{J}$-MEC apply for $\mathcal{J}$-augmented MAGs.

$\mathcal{I}$-Augmented MAGs

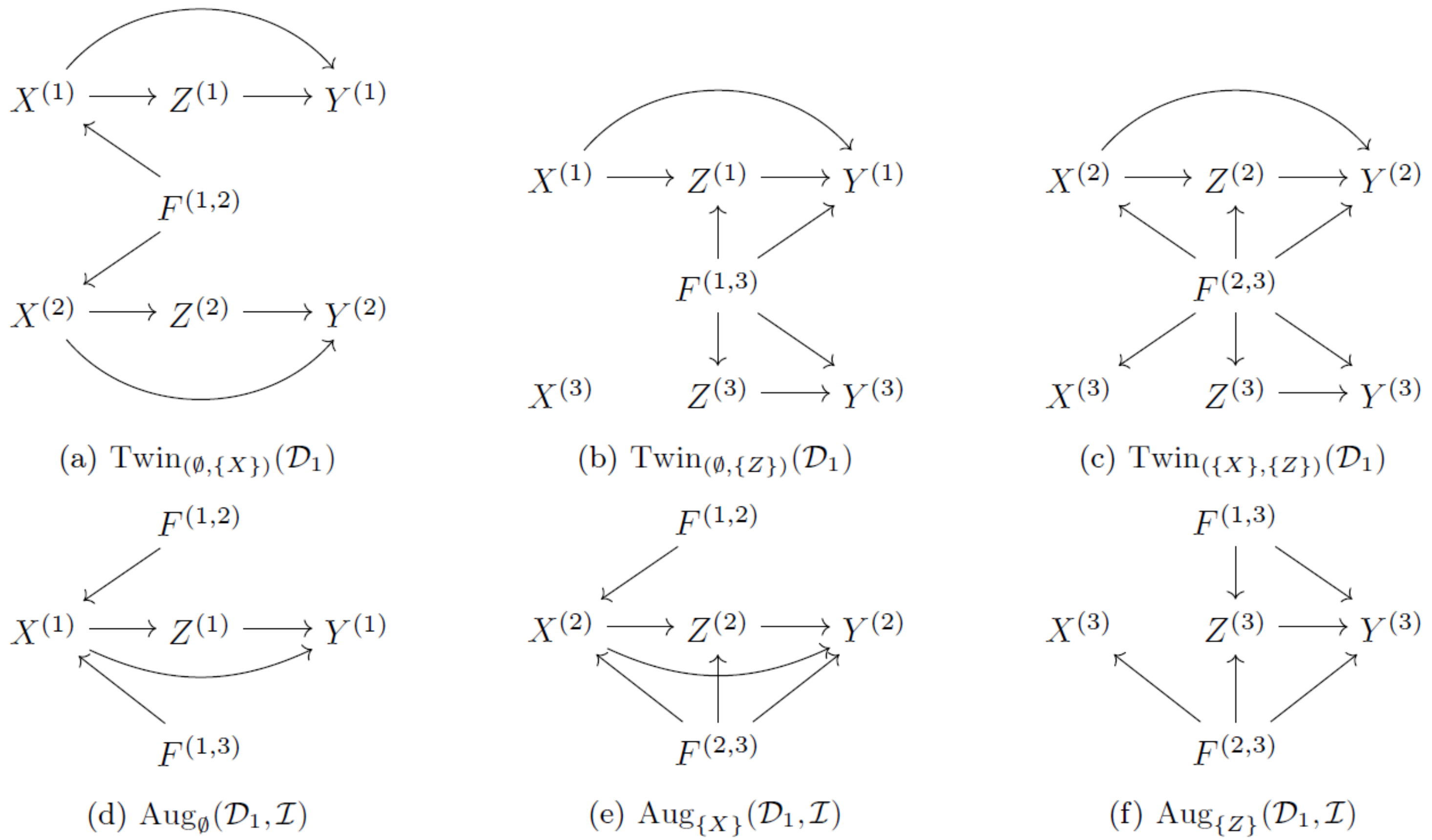


Figure 2: Illustration of the construction of $\mathcal{I}-$augmented MAGs from twin augmented MAGs. Figure 1a is the ground truth graph. The intervention targets are $\mathcal{I} = \{\mathbf{I}_1 = \emptyset, \mathbf{I}_2 = \{X\}, \mathbf{I}_3 = \{Z\}\}$. (a), (b), and (c) are the twin augmented MAGs. (d), (e), and (f) are the $\mathcal{I}$-augmented MAGs.

Learning by Combining Experiments

- The union graph of all $\mathcal{I}$-Markov equivalent $\mathcal{I}$-augmented MAGs marks what is fundamentally learnable.

Learning by Combining Experiments

- The union graph of all $\mathcal{I}$-Markov equivalent $\mathcal{I}$-augmented MAGs marks what is fundamentally learnable.
- Step 1: Learn the skeleton for each domain.

Learning by Combining Experiments

- The union graph of all $\mathcal{J}$-Markov equivalent $\mathcal{J}$-augmented MAGs marks what is fundamentally learnable.
- Step 1: Learn the skeleton for each domain.
- Step 2: Orient edges outwards from F nodes and apply the FCI rules.

Learning by Combining Experiments

- The union graph of all $\mathcal{J}$-Markov equivalent $\mathcal{J}$-augmented MAGs marks what is fundamentally learnable.
- Step 1: Learn the skeleton for each domain.
- Step 2: Orient edges outwards from F nodes and apply the FCI rules.
- Step 3: Apply the extra orientation rules.

Learning by Combining Experiments

- The union graph of all $\mathcal{J}$-Markov equivalent $\mathcal{J}$-augmented MAGs marks what is fundamentally learnable.
- Step 1: Learn the skeleton for each domain.
- Step 2: Orient edges outwards from F nodes and apply the FCI rules.
- Step 3: Apply the extra orientation rules.

Rule 9: For any $\mathbf{I} \in \mathcal{I}$, if $X \in \mathbf{I}$, $X^{(\mathbf{I})}, Y^{(\mathbf{I})}$ are adjacent in $\mathcal{G}_{\mathbf{I}}$, then orient $X^{(\mathbf{I})} o\!\!-\!\!* \, Y^{(\mathbf{I})}$ as $X^{(\mathbf{I})} \rightarrow Y^{(\mathbf{I})}$.
Rule 10: If $X^{(\mathbf{I})} \rightarrow Y^{(\mathbf{I})}$ in $\mathcal{G}_{\mathbf{I}}$ for some $\mathbf{I} \in \mathcal{I}$, replace the circle mark at $Y^{(\mathbf{J})}$ between $X^{(\mathbf{J})}$ and $Y^{(\mathbf{J})}$ in $\mathcal{G}_{\mathbf{J}}$ with an arrowhead for any $\mathbf{J} \in \mathcal{I} \setminus \{\mathbf{I}\}$.
Rule 11: In $\mathcal{G}_{\mathbf{I}}$, $\mathbf{I}, \mathbf{J} \in \mathcal{I}$, if $\mathbf{J} = \mathbf{I} \cup \{X\}$, $F^{(\mathbf{I},\mathbf{J})}$ is adjacent to $Y^{(\mathbf{I})}, Y \notin \mathbf{J}$, then orient $X^{(\mathbf{I})} *\!\!-\!\!* \, Y^{(\mathbf{I})}$ as $X^{(\mathbf{I})} \rightarrow Y^{(\mathbf{I})}$.

Experiments

- We compare the $\mathcal{I}$-MEC size under hard and soft interventions.

n	Mean under Hard	Mean under Soft	Graph	Ratio	Total Number of ADMGs
2	2.03 ± 0.15	2.93 ± 0.29	Random	0.69 ± 0.05	6
2	2.37 ± 0.12	3.67 ± 0.22	Complete	0.65 ± 0.05	6
3	19.50 ± 3.41	30.57 ± 4.36	Random	0.64 ± 0.11	200
3	14.03 ± 2.69	24.70 ± 4.12	Complete	0.57 ± 0.05	200
4	677.13 ± 227.72	1218.83 ± 361.83	Random	0.56 ± 0.18	34,752
4	721.37 ± 276.36	1529.57 ± 368.68	Complete	0.47 ± 0.07	34,752

Table 1: Estimation of $\mathcal{I}$-MEC size by enumerating all ADMGs of the same size. We consider random ADMGs and ADMGs by adding bidirected edges to random complete DAGs. For each setting we sample 30 ground truth ADMGs and calculate the mean and standard error of $\mathcal{I}$-MEC size and the ratio.

Summary

- We introduce generalized do-calculus rules for hard-interventional distributions.
- We propose a new graphical structure to characterize $\mathcal{I}$-MEC with hard interventions and latents.
- We design a sound learning algorithm that combines data from hard interventions.
- We conduct experiments to show that the hard $\mathcal{I}$-MEC size is on average smaller than soft $\mathcal{I}$-MEC.